

A system for organizing multi-dimensional pattern data into a reduceddimension representation comprising: a neural network comprised of a plurality of layers of nodes, the plurality of layers including: an input layer comprised of a plurality of input nodes, a hidden layer, and an output layer comprised of a plurality of non-linear output nodes, wherein the number of non-linear output nodes is less than the number of input nodes; receiving means for receiving multi-dimensional pattern data into the input layer of the neural network; output means for generating an output signal for each of the output nodes of the output layer of the neural network corresponding to received multi-dimensional pattern data; and training means for completing a training of the neural network, wherein the training means includes means for equalizing and orthogonalizing the output signals of the output nodes by reducing a covariance matrix of the output signals to the form of a diagonal matrix.

- 2. A system according to claim 1, wherein said training means uses backpropagation to iteratively update weights for the links between nodes of adjacent layers.
- 3. A system according to claim 2, wherein said weights are generated randomly in the interval (W, -W).
- 4. A system according to claim 3, wherein averaged variance of all dimensions of the multi-dimensional pattern data is:

$$V_{in} = \frac{1}{SP} \sum_{\substack{i=1 \ i=1}}^{S} \sum_{p=1}^{P} (x_{ip} - \langle x_i \rangle)^2$$

, and the elements of the covariance matrix of the output signals of the output nodes are defined

by:

$$V_{out,k_1k_2} = \frac{1}{P} \sum_{p=1}^{P} \left(O_{k_1p} - \langle O_{k_1} \rangle \right) \left(O_{k_2p} - \langle O_{k_2} \rangle \right)$$

,where p = 1, 2, ..., P;

 $O_{k_{i}p}$ is the output signal of the k_{1} th node of the output layer for the pth input data pattern vector;

 O_{k_2p} is the output signal of the k_2 th node of the output layer for the pth input data pattern vector;

 $\langle O_{k_1} \rangle$ is the average of O_{k_1p} evaluated over the set of input data pattern vectors

 $\langle O_{k_1} \rangle$ is the average of O_{k_1p} evaluated over the set of input data pattern vectors

$$k_1 = 1$$
 to K;

$$k_2 = 1$$
 to K;

K is the number of dimensions in the reduced-dimension representation; and

< > denotes the mean evaluated over the set of input data pattern vectors for each indicated component.

5. A system according to claim 4, wherein weights Δw_{kj} between the hidden layer and the output layer are iteratively updated according to the expression:

, where η is a constant of suitable value chosen to provide efficient convergence but to avoid oscillation;

O_{jp} is the output signal from the jth node in the layer preceeding the output layer due to the pth input data pattern vector;

E is the error given by:

$$E = \sum_{k_1=1}^{K} \sum_{k_2=k_1}^{K} E_{k_1 k_2}$$

and,

$$\stackrel{\frown}{E}_{k_1 k_2} = \left(\frac{V_{out,kk} - r_{kk} V_{in}}{r_{kk} V_{in}} \right)^2$$

, where $k_1 = k_2 = k$; k = 1,...,K; and r_{ik} is a positive constant which has an effect of increasing the speed of training,

$$E_{k_1 k_2} = \left(\frac{V_{out, k_1 k_2}}{r_{k_1 k_2}} V_{in} \right)^2$$

, where $k_2 > k_1$; $k_1 = 1, ..., K-1$; $k_2 = k_1 + 1, ..., K$; and $r_{k_1 k_2}$ is a positive constant which has an effect of increasing the speed of training; and

 $\delta_{kp} = \delta_{kp,1} + \delta_{kp,2} + \delta_{kp,3} \text{ , where } \delta_{kp} \text{ is a value proportional to the contribution to the}$ error E by the outputs of the kth node of the output layer, for the pth input data pattern vector, and $\delta_{kp,1}$, $\delta_{kp,2}$, and $\delta_{kp,3}$ are components of δ_{kp} .

6. A system according to claim 5, wherein:

$$\Delta w_{kj,1} = -\eta \frac{\partial E_{kk}}{\partial w_{kj}} = \frac{1}{P} \sum_{p=1}^{P} \eta \delta_{kp,1} O_{jp}$$

$$\Delta w_{kj,2} = -\eta \sum_{k_2 = k+1}^{j} \frac{\partial E_{kk_2}}{\partial w_{kj}} = \frac{1}{P} \sum_{p=1}^{P} \eta \delta_{kp,2} O_{jp}$$

$$\Delta w_{kj,3} = -\eta \sum_{k_1=1}^{k-1} \frac{\partial E_{k_1 k}}{\partial w_{kj}} = \frac{1}{P} \sum_{p=1}^{P} \eta \delta_{kp,3} O_{jp}$$

where $\Delta w_{kj,1}$ is the contribution from the diagonal terms of the covariance matrix of the outputs, $\Delta w_{kj,2}$ is the contribution from the off-diagonal terms in kth row,

 $\Delta w_{kj,3}$ is the contribution from the off-diagonal terms in kth column, and

 O_{jp} is the output signal from the jth node in the layer preceeding the output layer for the pth input data pattern vector.

7. A system according to claim 6, wherein:

$$\delta_{kp,1} = 4 (V_{out,kk} - r_{kk}V_{in})(\langle O_k \rangle - O_{kp})O_{kp}(1 - O_{kp})$$

$$\delta_{kp,1} = 2 \left(\sum_{k_2=k+1}^{K} V_{out,kk_2}(-O_{kp}) \right) O_{kp}(1-O_{kp})$$

$$\delta_{kp,3} = 2 \left(\sum_{k_1=1}^{k-1} \left(V_{out,k_1k} (< O_k > -O_{kp}) \right) O_{kp} (1 - O_{kp}) \right)$$

where

 O_{kp} is the output signal from the kth node in the output layer for the pth input data pattern vector, and

 $\langle O_{kp} \rangle$ is the average of O_{kp} evaluated over the set of input data pattern vectors.

8. A system according to claim 5, wherein backpropagation of error to the weights Δw_{ji} between the jth node in a layer of nodes and the ith node in its' preceeding layer:

$$\Delta w_{ji} = \eta \frac{\partial E}{\partial w_{ji}} = \frac{1}{P} \sum_{p=1}^{P} \eta \delta_{jp} x_{ip}$$

where, δ_{jp} is given by:

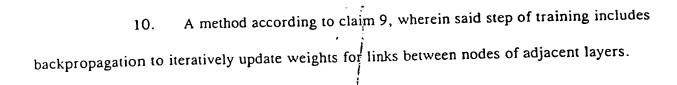
$$\delta_{jp} = \left(\sum_{k=1}^{K} \delta_{kp} w_{kj}\right) O_{jp} (1 - O_{jp})$$

9. A method for effecting the organization of multi-dimensional pattern data into a reduced dimensional representation using a neural network having an input layer comprised of a plurality of input nodes, a hidden layer, and an output layer comprised of a plurality of non-linear output nodes, wherein the number of non-linear output nodes is less than the number of input nodes, said method comprising:

receiving multi-dimensional pattern data into the input layer of the neural network;

generating an output signal for each of the ouput nodes of the neural network corresponding to received multi-dimensional pattern data; and

training the neural network by equalizing and orthogonalizing the output signals of the output nodes by reducing a covariance matrix of the output signals to the form of a diagonal matrix.



- 11. A method according to claim 10, wherein said weights are generated randomly in the interval (W, -W).
- 12. A method according to claim 11, wherein averaged variance of all dimensions of the multi-dimensional pattern data is:

$$V_{in} = \frac{1}{SP} \sum_{i=1}^{S} \sum_{p=1}^{P} (x_{ip} - \langle x_i \rangle)^2$$

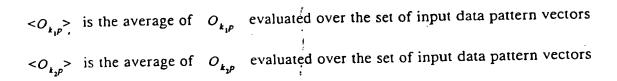
, and the elements of the covariance matrix of the output signals of the output nodes is:

$$V_{out,k_1k_2} = \frac{1}{P} \sum_{p=1}^{P} \left(O_{k_1p} - \langle O_{k_1} \rangle \right) \left(O_{k_2p} - \langle O_{k_2} \rangle \right)$$

,where p = 1, 2, ..., P;

 O_{k_1p} is the output signal of the k_1 th node of the output layer for the pth input data pattern vector;

 $O_{k_{2}p}$ is the output signal of the k_{2} th node of the output layer for the pth input data pattern vector;



$$k_1 = 1$$
 to K;

$$k_2 = 1$$
 to K ;

K is the number of dimensions in the reduced-dimension representation; and

< > denotes the mean evaluated over the set of input data pattern vectors for each indicated component.

13. A method according to claim 12, wherein weights Δw_{kj} between the hidden layer and the output layer are iteratively updated according to the expression:

$$\Delta w_{kj} = -\eta \frac{\partial E}{\partial w_{kj}} = \frac{1}{P} \sum_{p=1}^{P} \eta \delta_{kp} O_{jp}$$

$$= -\eta \left(\frac{\partial E_{kk}}{\partial w_{kj}} + \sum_{k_2=k+1}^{K} \frac{\partial E_{kk_2}}{\partial w_{kj}} + \sum_{k_1=1}^{k-1} \frac{\partial E_{k_1k}}{\partial w_{kj}} \right)$$

$$= \Delta w_{kj,1} + \Delta w_{kj,2} + \Delta w_{kj,3}$$

, where η is a constant of suitable value chosen to provide efficient convergence but to avoid oscillation;

 O_{jp} is the output signal from the jth node in the layer preceeding the output layer, due to the pth input data pattern vector;

E is the error given by:

$$E = \sum_{k_1 = 1 \atop k_1 = 1}^{K_1} \sum_{k_2 = k_1}^{K} E_{k_1 k_2}$$

and,

$$\widetilde{E_{k_1 k_2}} = \left(\frac{V_{out,kk} - r_{kk} V_{in}}{r_{kk} V_{in}} \right)^2$$

,where $k_1 = k_2 = k$; k = 1,...,K; and k_k is a positive constant which has an effect of increasing the speed of training,

$$E_{k_1 k_2} = \left(\frac{V_{out, k_1 k_2}}{r_{k_1 k_2}} V_{in} \right)^2$$

, where $k_2 > k_1$; $k_1 = 1, ..., K-1$; $k_2 = k_1 + 1, ..., K$; and $r_{k_1 k_2}$ is a positive constant which has an

effect of increasing the speed of training; and

 $\delta_{kp} = \delta_{kp,1} + \delta_{kp,2} + \delta_{kp,3} \text{ , where } \delta_{kp} \text{ is a value proportional to the contribution to the}$ error E by the outputs of the kth node of the output layer, for the pth input data pattern vector, and $\delta_{kp,1}$, $\delta_{kp,2}$, and $\delta_{kp,3}$ are components of δ_{kp}

14. A method according to claim 13, wherein:

$$\Delta w_{kj,1} = -\eta \frac{\partial E_{kk}}{\partial w_{kj}} = \frac{1}{P} \sum_{p=1}^{P} \eta \delta_{kp,1} O_{jp}$$

$$\Delta w_{kj,2} = -\eta \sum_{k_2=k+1}^{K} \frac{\partial E_{kk_2}}{\partial w_{kj}} = \frac{1}{P} \sum_{p=1}^{P} \eta \delta_{kp,2} O_{jp}$$

$$\Delta w_{kj,3} = -\eta \sum_{k_1=1}^{k_1-1} \frac{\partial E_{k_1 k}}{\partial w_{kj}} = \frac{1}{P} \sum_{p=1}^{P} \eta \delta_{kp,3} O_{jp}$$

where $\Delta w_{kj,l}$ is the contribution from the diagonal term,

 $\Delta W_{kj,2}$ is the contribution from the off-diagonal terms in kth row, and $\Delta W_{kj,3}$ is the contribution from the off-diagonal terms in kth column.

$$\delta_{kp,1} = 4(V_{out,kk} - r_{kk}V_{in})(\langle O_k \rangle - O_{kp})O_{kp}(1 - O_{kp})$$

$$\delta_{kp,1} = 2 \left(\sum_{k_2=k+1}^{K} V_{out,k_2}(\langle O_k \rangle - O_{kp}) \right) O_{kp}(1 - O_{kp})$$

$$\delta_{kp,3} = 2 \left(\sum_{k_1=1}^{k-1} V_{out,k_1k}(-O_{kp}) \right) O_{kp}(1-O_{kp})$$

, where

 O_{kp} is the output signal from the kth node in the layer preceeding the output layer for the pth input data pattern vector, and

<O_{kp}> is the average of O_{kp} evaluated over the set of input data pattern vectors.

16. A method according to claim 13, wherein backpropagation of error to the weights Δw_{ji} between the jth node in a layer of nodes and the ith node in its' preceeding layer are:

$$\Delta w_{ji} = \eta \frac{\partial E}{\partial w_{ji}} = \frac{1}{P} \sum_{p=1}^{P} \eta \delta_{jp} x_{ip}$$

where, δ_{jp} is given by:

$$\delta_{jp} = \left(\sum_{k=1}^{K} \delta_{kp} w \Big|_{kj}\right) O_{jp} (1 - O_{jp})$$

17. A system for organizing multi-dimensional pattern data into a reduced dimensional representation comprising:

a neural network comprised of a plurality of layers of nodes, the

plurality of layers including:

an input layer comprised of a plurality of input nodes, and an output layer comprised of a plurality of non-linear output

nodes, wherein the number of non-linear output nodes is less than the number of input nodes;

receiving means for receiving multi-dimensional pattern data into the

input layer of the neural network;

output means for generating an output signal at the output layer of the neural network corresponding to received multi-dimensional pattern data; and

training means for completing a training of the neural network, wherein the training means conserves a measure of the total variance of the output nodes, wherein the total variance of the output nodes is defined as:

$$V = (1/P) \sum_{p=1}^{p=P} \sum_{i=1}^{i=S} (x_{ip} - \langle x_i \rangle)^2$$

,where $\{x_p\}$ is a set of data pattern vectors;

p = 1, 2, ..., P;

P is defined as a positive integer;

 $< x_i >$ denotes the mean value of of x_{ip} evaluated over the set of data pattern

vectors;

S is the number of dimensions;

 x_{ip} is the ith component of x_p , the pth member of a set of data pattern vectors.

18. A system according to claim 17, wherein said training means completes the training of the neural network via backpropagation for progressively changing weights for the output nodes.

19. A system according to claim 18, wherein said training means further includes,

means for training the neural network by backpropagation by progressively changing weights w_{kj} at the output layer of the neural network in accordance with,

$$\Delta w_{kj} = (1/P) \sum_{p=1}^{p=P} \Delta w_{p,kj} = (1/P) \sum_{p=1}^{p=P} \eta \delta_{pk} O_{pj}$$

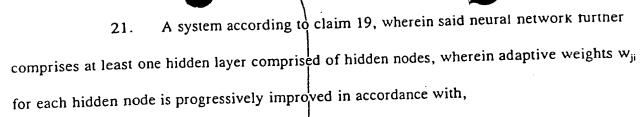
, where O_{pj} is the output signal from the jth node in the layer preceeding the output layer due to the pth data pattern,

 η is a constant of suitable value chosen to provide efficient convergence but to avoid oscillation, and

 δ_{pk} is a value proportional to the contribution to the error E by the outputs of the kth node of the output layer for the pth input data pattern.

20. A system according to claim 19, wherein:

$$\delta_{pk} = [V - (1/P) \sum_{q} \sum_{n} (O_{qn} - \langle O_n \rangle^2] (O_{pk} - \langle O_k \rangle) O_{pk} (1 - O_{pk})$$



$$\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}} = \frac{1}{P} \sum_{p=1}^{p=P} \eta \delta_{pj} O_{pi}$$

, where O_{pi} is the output signal for the ith node of the layer preceeding the jth layer of the pth input data pattern.

22. A system according to claim 21, wherein:

$$\delta_{pj} = \left(\sum_{k=1}^{K} \delta_{pk}^{k} w_{kj}\right) O_{pj} (1 - O_{pj})$$

data into a reduced dimensional representation using a neural network having an input layer comprised of a plurality of input nodes, and an output layer comprised of a plurality of non-linear output nodes, wherein the number of non-linear output nodes are less than the number of input nodes, said method comprising:

receiving a set $\{x_p\}$ of data pattern vectors into the input layer of the neural network, wherein p=1,2,...,P and wherein P is defined as a positive integer, and wherein the set of data pattern vectors has a total variance defined as,

$$V = (1/P) \sum_{p=1}^{p=P} \left| \sum_{i=1}^{i=S} (x_{ip} - \langle x_i \rangle)^2 \right|$$

, where $\{x_p\}$ is a set of data pattern vectors,

$$p = 1, 2, ..., P;$$

P is defined as a positive integer;

 $< x_i >$ denotes the mean value of of x_{ip} evaluated over the set of data pattern vectors;

S is the number of dimensions;

 x_{ip} is the ith component of x_p , the pth member of a set of data pattern vectors;

training the neural network by backpropagation; and

displaying a multi-dimensional output signal from the output layer of the

neural network.





24. A method according to claim 23, wherein said step of training the neural network by backpropagation includes progressively changing weights w_{kj} at the output layer of the neural network in accordance with,

$$\Delta w_{kj} = (1/P) \sum_{p=1}^{p=P} \Delta w_{p,kj} = (1/P) \sum_{p=1}^{p=P} \eta \delta_{pk} O_{pj}$$

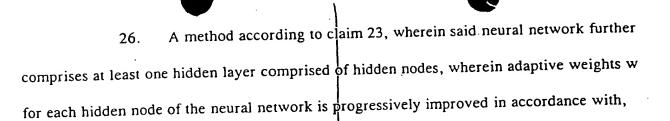
, where O_{pj} is the output signal from the jth node in the layer preceeding the output layer due to the pth data pattern, and

 η is a constant of suitable valure chosen to provide efficient convergence but to avoid oscillation.

 δ_{pk} is a value proportional to the contribution to the error E by the outputs of the kth node of the output layer for the pth input data pattern.

25. A system according to claim 24, wherein:

$$\delta_{pk} = [V - (1/P) \sum_{q} \sum_{n} (O_{qn}^{-} - \langle O_{n}^{-} \rangle^{2}] (O_{pk}^{-} - \langle O_{k}^{-} \rangle) O_{pk} (1 - O_{pk})$$



$$\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}} = \frac{1}{P} \sum_{p=1}^{p=P} \eta \delta_{pj} O_{pi}$$

, where O_{pi} is the output signal for the ith node of the layer preceeding the jth layer of the pth input data pattern.

27. A method according to claim 26, wherein

$$\delta_{pj} = \left(\sum_{k=1}^{K} \delta_{pk} w_{kj}\right) O_{pj} (1 - O_{pj})$$

- 28. A method according to claim 23, wherein said multi-dimensional output signal is a two-dimensional output signal.
- 29. A method according to claim 23, wherein said two-dimensional output signal includes data points plotting in relation to 2-dimensional axes.

add